Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2018, vol.2018-January, pages 2843-2852

Quantifying volume, velocity, and variety to support (Big) data-intensive application development

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2017 IEEE. In the era of digital economies, data can be considered as the new commodity, fueling the next-generation software services and applications. Increasing amounts of data, generated on a daily basis by various domains, such as social networks, stock exchanges, the Internet of Things, and cyber-physical systems, are soon expected to exceed the yottabyte1 frontier. To process this overwhelming amount, Big Data solutions are being developed to enable a new generation of data-centric/data-intensive applications (DIAs) and services. However, many of such applications currently fail to meet the increasingly demanding data management requirements. In particular, proper techniques and tools to support architects and developers in DIA design are required to cope with these pressing Big Data challenges. This paper makes an initial step in this direction, aiming at reducing the gap between the architects and DIAs they have to develop. The proposed approach extends the conventional Big Data process workflow with a way of capturing and modeling the 'three Vs' of Big Data (i.e. volume, velocity, and variety) to provide useful insights on the overall process, knowing the behavior of its individual components. Starting from the V-attributes of the Big Data process components, the proposed framework provides an estimation of its V-metrics by evaluating a performance model generated from the process. To demonstrate the feasibility and the effectiveness of the approach, a case study on a computer vision DIA is reported.

http://dx.doi.org/10.1109/BigData.2017.8258252

Keywords

Big Data, Queuing Networks, Variety, Velocity, Volume, Workflow

References

[1] Balsamo, S., Marzolla, M. : Performance Evaluation of UML Software Architectures with Multiclass Queueing Network Models. In: Proceedings of the 5th International Workshop on Software and Performance. pp. 37-42. ACM (2005)

[2] Barbierato, E., Gribaudo, M., Iacono, M. : Performance evaluation of NoSQL big-data applications using multi-formalism models. Future Generation Computer Systems 37, 345-353 (2014)

[3] Bertoli, M., Casale, G., Serazzi, G. : JMT: performance engineering tools for system modeling. SIGMETRICS Perform. Eval. Rev. 36 (4), 10-15 (2009)

[4] Boehm, B. W., Brown, J. R., Lipow, M. : Quantitative Evaluation of Software Quality. In: Proceedings of the 2nd International Conference on Software Engineering. pp. 592-605. IEEE Computer Society Press (1976)

[5] Bruneo, D., Distefano, S., Longo, F., Scarpa, M. : Stochastic Evaluation of QoS in Service-Based Systems. IEEE Trans. Parallel Distrib. Syst. 24 (10), 2090-2099 (2013)

[6] Cardoso, J., Sheth, A., Miller, J., Arnold, J., Kochut, K. : Quality of service for workflows and web service processes. Web Semantics: Science, Services and Agents on the World Wide Web 1 (3), 281-308 (2004)

[7] Chen, C. P., Zhang, C. Y. : Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 275, 314-347 (2014)

[8] Chen, M., Mao, S., Liu, Y. : Big Data: A Survey. Mobile networks and applications 19 (2), 171-209 (2014)

[9] Cisco R- : The zettabyte era: Trends and analysis. Tech. rep., Cisco R- Visual Networking Index (2017), https://www. cisco. com/c/en/us/ solutions/collateral/service-provider/visual-networking-index-vni/ vni-hyperconnectivity-wp. html, accessed on 20/09/2017

[10] Cox, M., Ellsworth, D. : Application-controlled demand paging for out-of-core visualization. In: Proceedings of the 8th conference on Visualization'97. IEEE Computer Society Press (1997)

[11] Dautov, R., Distefano, S., Merlino, G., Puliafito, A., Longo, F. : Towards a Global Intelligent Surveillance System. In: Proceedings of the 11th International Conference on Distributed Smart Cameras (ICDSC 2017). ACM (2017)

[12] Dautov, R., Paraskakis, I., Stannett, M. : Utilising stream reasoning techniques to underpin an autonomous framework for cloud application platforms. Journal of Cloud Computing 3 (1), 13 (2014)

[13] Dautov, R., Stannett, M., Paraskakis, I. : On the role of stream reasoning in run-time monitoring and analysis in autonomic systems. In: Proceedings of the 8th South East European Doctoral Student Conference (DSC 2013). SEERC (2013)

[14] Dean, J., Ghemawat, S. : MapReduce: Simplified Data Processing on Large Clusters. Commun. ACM 51 (1), 107-113 (2008)

[15] Gani, A., Siddiqa, A., Shamshirband, S., Hanum, F. : A survey on indexing techniques for big data: Taxonomy and performance evaluation. Knowledge and Information Systems 46 (2), 241-284 (2016)

[16] Gantz, J., Reinsel, D. : Extracting value from chaos. IDC iView 1142, 1-12 (2011)

[17] Gérard, S., Selic, B. : The UML-MARTE Standardized Profile. vol. 41, pp. 6909-6913. Elsevier (2008)

[18] IBM Corporation: Four Vs of Big Data (2015), http: //www. ibmbigdatahub. com/infographic/four-vs-big-data/, accessed on 20/09/2017

[19] Lazowska, E. D., Zahorjan, J., Graham, G. S., Sevcik, K. C. : Quantitative System Performance: Computer System Analysis Using Queueing Network Models. Prentice-Hall, Inc. (1984)

[20] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. : Big data: The next frontier for innovation, competition, and productivity. Tech. rep., McKinsey Global Institute (2011)

[21] McAfee, A., Brynjolfsson, E. : Big data: The management revolution. Harvard business review 90 (10), 60-66 (2012)

[22] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., Belfkih, S. : Big Data Technologies: A Survey. Journal of King Saud University-Computer and Information Sciences (2017)

[23] Shen, C., Tong, W., Hwang, J. N., Gao, Q. : Performance Modeling of Big Data Applications in the Cloud Centers. J. Supercomput. 73 (5), 2258-2283 (2017)

[24] Singh, D., Reddy, C. K. : A Survey on Platforms for Big Data Analytics. Journal of Big Data 2 (1) (2015)

[25] Wang, L., Zhan, J., Luo, C., Zhu, Y., Yang, Q., He, Y., Gao, W., Jia, Z., Shi, Y., Zhang, S., et al. : Bigdatabench: A big data benchmark suite from Internet services. In: 2014 IEEE 20th International Symposium on High Performance Computer Architecture (HPCA). pp. 488-499. IEEE (2014)

[26] Wimmer, M., Albutiu, M. C., Kemper, A. : Optimized workflow authorization in service oriented architectures. In: Emerging Trends in Information and Communication Security, pp. 30-44. Springer (2006)

[27] Woodside, M., Petriu, D. C., Petriu, D. B., Shen, H., Israr, T., Merseguer, J. : Performance by unified model analysis (PUMA). In: Proceedings of the 5th international workshop on Software and performance. pp. 1-12. ACM (2005)

[28] Wu, X., Woodside, M. : Performance modeling from software components. In: ACM SIGSOFT Software Engineering Notes. vol. 29, pp. 290-301. ACM (2004)

[29] YouTube: YouTube Official Blog (2017), https://youtube. googleblog. com/2017/08/an-update-on--ur-commitment-to-fight. html, accessed on 20/09/2017